Model building of disulfide bonds in proteins with known three-dimensional structure

Hazes, Bart; Dijkstra, Bauke W.

Published in:
%22Protein Engineering%2C Design and Selection%22

DOI:
10.1093/protein/2.2.119

Protein Engineering vol.2 no.2 pp.119–125, 1988

Model building of disulfide bonds in proteins with known three-dimensional structure

Bart Hazes and Bauke W.Dijkstra[1]

Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

[1]To whom correspondence should be addressed

As an aid in the selection of sites in a protein where a disulfide bond might be engineered, a computer program has been developed. The algorithm starts with the generation of Cβ positions from the N, Cα and C atom coordinates available from a three-dimensional model. A first set of residue pairs that might form a disulfide bond is selected on the basis of Cβ–Cβ distances between residues. Then, for each residue in this set, Sγ positions are generated, which satisfy the requirement that, with ideal values for the Cα–Cβ and Cβ–Sγ bond lengths and for the bond angle at Cβ, the distance between Sγ of residue 1 and Cβ of residue 2 in a pair (determined by the bond angle at Sγ2) is at, or very close to its ideal value. Usually two acceptable Sγ positions are found for each half cystine, resulting in up to four different conformations for the disulfide bond. Finally, these conformations are subjected to an energy minimization procedure to remove large deviations from ideal geometry and their final energies are calculated. User input determines which final conformations are energetically acceptable. These conformations are written to a file to allow further analysis and e.g. inspection on a computer graphics device.

Key words: computer modelling/disulfide bonds/model building/prediction/SS-bonds

Introduction

In recent years protein engineering has become a powerful experimental tool to modify at will the properties of polypeptide molecules. One of the goals in the use of this technique has been to increase the stability of enzymes in order to broaden their utility in commercial and medical applications. Disulfide bonds may serve to stabilize the native conformation of proteins (Wetzel, 1987). The detailed mechanism by which these bonds confer stability is not known but presumably they lower the entropy of the unfolded form (Anfinsen and Scheraga, 1975). Several investigators have analyzed the stabilizing effect of disulfide bonds by comparing the stability of native proteins with that of mutants into which a disulfide bond had been engineered. The results are not unequivocal.

Perry and Wetzel (1984) introduced a disulfide bond into T4 lysozyme between residues 3 and 97. Provided the intrinsic Cys 54 residue has a blocked thiol group, then the mutant enzyme with an intact 3–97 disulfide bond was found to have enhanced temperature stability. The reduced form behaved essentially identical to wild-type T4 lysozyme. However, when the thiol group of residue 54 was not blocked, the mutant showed no enhanced thermal stability, neither in the oxidized nor in the reduced form.

Also in dihydrofolate reductase from *Escherichia coli* a disulfide bond has been engineered (Villafranca *et al.*, 1983, 1987). The oxidized (cross-linked) form of this enzyme was not more resistant to thermal denaturation than the wild-type enzyme. On the other hand, the disulfide-cross-linked enzyme was shown to be more stable with respect to unfolding as measured by guanidine hydrochloride denaturation.

Two reports exist on the stabilization upon disulfide bond formation in subtilisin BPN'. Pantoliano *et al.* (1987) found an enhanced stabilization against thermal inactivation after engineering a disulfide bond between residues 22 and 87. To reduce the amount of autolysis the stability of the enzyme and mutant was determined in the presence of an inhibitor. Wells and Powers (1986) produced the same mutant. These authors were interested primarily in increasing the stability against autolysis. Their results indicate, however, that a disulfide bond between residues 22 and 87 actually lowers the autolytic stability. A disulfide bond introduced between residues 24 and 87 by these authors did not affect stability against autolysis.

In two of the above-mentioned cases the three-dimensional structures of the disulfide mutants were analyzed by X-ray crystallography. The geometry of the disulfide bond in dihydrofolate reductase (Villafranca *et al.*, 1987) appeared not to be identical with that of any of the previously known protein disulfide bonds (Richardson, 1981). The Pro 39 to Cys mutation caused some small conformational changes in the region of the substitution only. No displacements of α-carbon atoms larger than 0.25 Å appeared anywhere in the molecule. Katz and Kossiakoff (1986) analyzed the X-ray structures of two subtilisin mutants. Also here it appeared that the engineered disulfide bonds adopted novel sets of dihedral angles, which did not fit well into any of the previously found geometric categories with respect to all five dihedral angeles. The main chain atoms in the disulfide region showed only minor differences.

It is clear that, in order to assess the role of disulfide bonds in the stabilization of protein conformation, more research is required. Given the atomic coordinates of a protein molecule, then model building is a first step to identify potential sites in a protein where a disulfide group of good stereochemistry can be inserted with a minimum of main chain rearrangement. A protein of 200 amino acid residues has in principle 20 000 possibilities for disulfide bond formation. Most of the amino acid residues of course will be too far apart to form these bonds, but for the other residues all possibilities giving disulfide bonds with the correct geometry should be taken into account. This task is most easily done by computer. Pabo and Suchanek (1986) have published an algorithm for locating sites for potential disulfide bonds in proteins of known three-dimensional structure. Their method encompasses the comparison of main chain atoms of two residues with the main chain atoms of two cysteine residues forming a disulfide bond of known conformation. If the backbone atoms of such a set have the same spatial relationship as the Cys-backbone atoms, then these residues in the set might provide a plausible position for introducing a disulfide. This means that in these authors' method, the conformation of disulfide bonds

which can be predicted is restricted to the conformations of disulfide bonds for which a three-dimensional structure has been determined. Pantoliano *et al.* (1987) used a similar approach to model disulfide bonds into proteins as Pabo and Suchanek (1986). From the few examples available at this time, however, it is clear that not all engineered disulfide bonds will have a conformation similar to the ones already known.

Here we present a quite different and also much faster method to predict conformations of disulfide bonds in proteins of known three-dimensional structure. Our method is not limited by the number of disulfides for which the three-dimensional structure has been solved by X-ray crystallography. For each amino acid residue with known positions of the main chain N, Cα and C atoms, the position of a Cβ atom and a starting position of an Sγ atom are generated (Sγ in the plane of N, Cα and Cβ). Selection of potential disulfide bond-forming residue pairs is done first on the basis of their Cβ−Cβ distance. Subsequently the Sγ positions of these selected residues are evaluated on the basis of their Cβ1−Sγ2 and Cβ2−Sγ1 distance, while rotating the Sγ atom about the Cα−Cβ bond of residue 2, respectively residue 1. To ensure a correct geometry for the atoms in the disulfide bond with respect to the bond lengths and bond angles, the atoms in the disulfide bond are subjected to an energy minimization procedure.

Materials and methods

Philosophy of the program

Before going on to describe the features and working of the program in some detail it may be useful to touch briefly upon the basic ideas of the program. Our main goal was to develop a computer program which can generate any potential disulfide bond within the explicit conformational and energy constraints provided by the user. Not a single conformation should be excluded *a priori* because of the limitations imposed by the finite size of a database with known disulfide conformations. The coordinates of the N, Cα and C atoms of an amino acid residue are sufficient to generate positions for a Cβ and Sγ atom. Therefore, this information alone should suffice to allow the program to generate all potential disulfide bonds. If more than one possible conformation for a disulfide bond is found, the program should rank these conformations according to the quality of their geometry. Finally, the program should be easy to use, with default values for input parameters where appropriate. It should also be written such that it can be run on a wide variety of computers without the need for modifying code specific for any one computer make.

General description of the algorithm

Figure 1 shows an outline of the program SSBOND in diagram form. The program has been written in standard FORTRAN 77 with no machine-specific extensions. It runs on computers as different as VAX 11/750, μVAX II, CDC Cyber 176 and the CONVEX C1-XP.

After initialization (subroutine SSINIT) the input file with atomic coordinates is read (subroutine RDFILE). The program accepts data in the Protein Data Bank format (Bernstein *et al.*, 1977) and data in the format of the BIOMOL set of protein structure determination programs. Other formats can easily be incorporated. Residues for which minimal coordinates for the N, Cα and C atoms are present are accepted. The coordinates of these atoms are stored, as well as the coordinates of Cβ atoms when available. These observed Cβ atom positions are used later for an analysis of the discrepancies between the ideal bond lengths

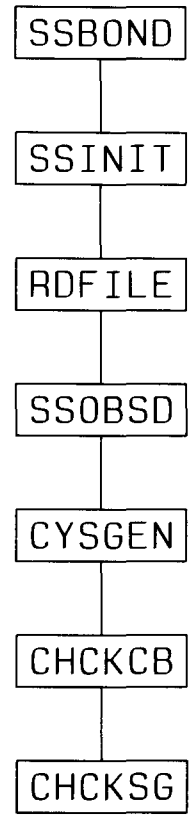

Fig. 1. Flow scheme of the SSBOND program. Indicated are the names of all major subroutines. See text for detailed description.

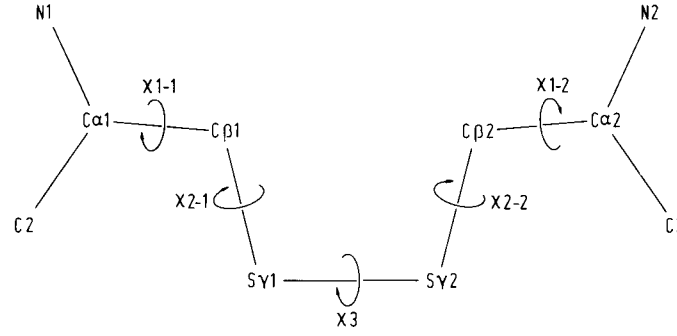

Fig. 2. Nomenclature of the dihedral angles in a disulfide bond.

and bond angles used in the program and the actual values occurring in the model. Also the names of the residues and their residue number are kept in memory. If cysteine residues are present, then the coordinates of the Sγ atoms are also stored. Subroutine SSOBSD uses these coordinates to analyze the conformations of already existing disulfide bonds in the protein.

The next step in the program is the generation of Cβ and Sγ positions for all residues. The Sγ position is calculated assuming the χ-1 angle is 0.0° (see Figure 2 for a definition of the dihedral angles in a disulfide). This is done in subroutine CYSGEN (see below for details). In addition, this subroutine calculates the root mean square (r.m.s.) difference between observed and calculated Cβ positions. This r.m.s. difference can be used in the selection of residue pairs which are close enough to form a disulfide bond.

Finally, subroutines CHCKCB and CHCKSG perform the selection of possible disulfide bond conformations. CHCKCB selects residue pairs which have their Cβ atoms at a distance

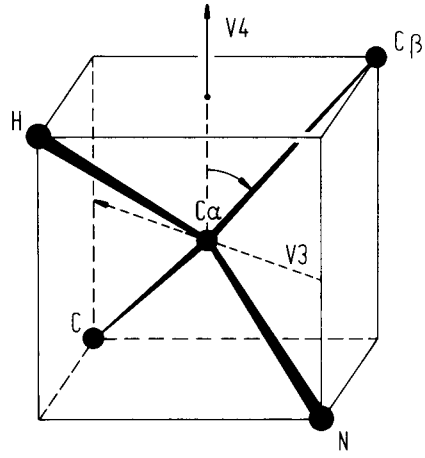

Fig. 3. Stereochemistry around the tetrahedral Cα atom of L-amino acid residues. All bond lengths in this figure have been normalized to 1.0 Å. *V3* is the normalized vector from the N to the C atom, positioned at the Cα atom as the origin. *V4* is −(the bisecting vector between the normalized Cα−N and Cα−C vectors). *V4* also has been normalized to 1.0 Å.

suitable for disulfide bond formation. CHCKSG determines the χ-1 angles at which good disulfide bonds can occur and restrains the bond lengths and bond angles in the predicted disulfide bond conformations to the ideal values by an energy minimization procedure. The energy of the predicted conformations is calculated and, optionally, the generated disulfide bonds are written to an output coordinate file.

Generation of Cβ positions

Figure 3 shows the stereochemistry around a tetrahedral Cα atom of an L-amino acid residue. All bond lengths in this figure are assumed to have been normalized to a length of 1.0 Å.

V3 is a normalized vector from N to C with the Cα atom as origin. Note that this vector is perpendicular to the plane of H, Cα and Cβ. The normalized vector $V4 = -(V1 + V2)$, in which *V1* and *V2* are normalized vectors from Cα to C, respectively Cα to N. *V4* lies in the plane of H, Cα and Cβ. Now it is easy to calculate the Cβ position:

$$x_{C\beta} = x_{C\alpha} + d_{CC} \times [R] \times V4$$

In this formula d_{CC} is the bond length between Cα and Cβ, and [*R*] is a rotation matrix which specifies a rotation about the vector *V3* [see e.g. Altman (1986) for a detailed description of such a rotation matrix]. In the case of ideal tetrahedral geometry at an α-carbon atom this rotation angle would have been half the perfect tetrahedral angle. However, the geometry at the α-carbon atom is known not to be perfectly tetrahedral. Therefore, and also to determine the optimal value for the Cα−Cβ bond length, we analyzed our database of refined high resolution structures (see Table II and below). In this database the mean Cα−Cβ distance is 1.538 ± 0.028 Å for all 6488 amino acid residues except glycine. For 241 cysteine residues this mean distance was found to be 1.532 ± 0.023 Å. This latter value was used throughout the subsequent analyses. Also the rotation angle was determined from this database. This angle appeared to be 52.3 ± 4.8° for all residues, and 53.2 ± 4.4° for the 241 Cys residues. We used the value 53.2°.

Generation of starting Sγ positions

The position of the Sγ atom in a cysteine residue is a function of χ-1, the torsion angle about the Cα−Cβ bond. For χ-1 = 0.0, Sγ lies in the plane of N, Cα and Cβ (Figure 4) and its position is then easy to calculate.

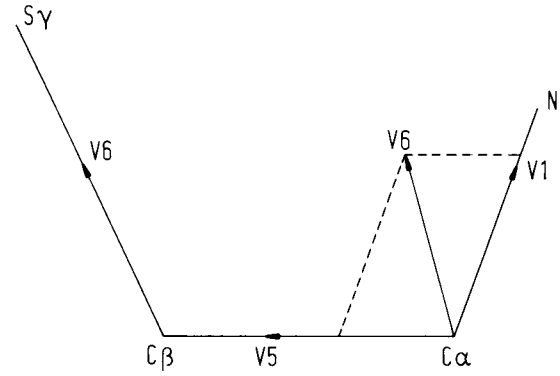

Fig. 4. Geometry of a cysteine residue with the χ-1 torsion angle + 0.0°. *V1*, *V5* and *V6* are normalized vectors along the bonds indicated.

$$x_{S\gamma} = x_{C\beta} + d_{CS} \times V6$$

d_{CS} is the bond length of the Cβ−Sγ bond. *V6* is given by (see Figure 4):

$$V6 = \text{norm6} \times (p \times V1 + q \times V5)$$

norm6 is a normalization factor; *V5* is given above; *V1* is the normalized vector from Cα to N. The constants *p* and *q* can easily be calculated from Figure 4:

$$p = \cos(\theta_{C\beta} - 90.0)/\cos(\theta_{C\alpha} - 90.0)$$
$$q = \sin(\theta_{C\beta} - 90.0) + p \times \sin(\theta_{C\alpha} - 90.0)$$

$\theta_{C\beta}$ is the standard bond angle Cα−Cβ−Sγ. $\theta_{C\alpha}$ is the calculated bond angle N−Cα−Cβ. From a survey of our database, including 95 disulfide bonds, the mean Cβ−Sγ bond length was found to be 1.81 Å. The N−Cα−Cβ and Cα−Cβ−Sγ bond angles were found to be 110.0 and 113.7°, respectively. The Sγ1−Sγ2 bond length was 2.03 Å and the Cβ1−Sγ1−Sγ2 bond angle was 104.0°. These values were used throughout the program.

For χ-1 values ≠ 0.0, *V6* is rotated over the desired χ-1 angle about vector *V5* using the rotation matrix [*R*] given above.

Selection of potential disulfide bond-forming residues

With standard bond lengths of 2.03 and 1.81 Å for the S−S and C−S bonds, respectively, and a standard value of 104° for the Cβ−Sγ1−Sγ2 bond angle, the Cβ1−Cβ2 distance in a disulfide bond can be calculated to vary between a minimum of 2.9 Å and a maximum of 4.6 Å. This distance is a function of the dihedral angle χ-3 (the Cβ1−Sγ1−Sγ2−Cβ2 dihedral angle). The minimum distance occurs at χ-3 = 0.0° and the maximum at χ-3 = 180.0°. However, in disulfide bonds found in proteins this χ-3 angle has a very strong preference for the values ±90° (Richardson, 1981). The Cβ1−Cβ2 distance corresponding with these values of χ-3 is 3.83 Å. Our first criterion to select acceptable disulfide bonds is based on this Cβ1−Cβ2 distance. Because of uncertainties in the observed atomic coordinates this criterion can not be applied rigorously. Instead we make this criterion less strict on the basis of an analysis of the discrepancies between the observed and calculated Cβ positions (see Results). In addition, an allowable deviation from the preferred χ-3 angles may be specified.

The second criterion for selecting acceptable disulfide bonds is the distance between the Cβ of one residue (Cβ1) and the Sγ atom of the other (Sγ2). With the same bond lengths and bond

Table I. Formulae and parameters used for the calculation of the energy of a disulfide bond

$E_{total} = E_{bond} + E_{bond\ angle} + E_{dihedral}$

$E_{bond} = \frac{1}{2}K\ (b-b_0)^2$; b is the actual bond length

	K(kcal/mol Å^2)	b_0 (Å)
Cα−Cβ bond	800	1.532
Cβ−Sγ bond	900	1.813
Sγ−Sγ bond	1000	2.030

$E_{bond\ angle} = \frac{1}{2}K\ (\theta-\theta_0)^2$; θ is the actual bond angle

	K(kcal/mol.rad^2)	θ_0 (°)
N−Cα−Cβ angle	110	109.5
C−Cα−Cβ angle	110	109.5
Cα−Cβ−Sγ angle	110	113.8
Cβ−Sγ−Sγ angle	110	104.0

$E_{dihedral} = K(1 + \cos n\chi)$; χ is the actual dihedral angle

	K(kcal/mol)	n
N−Cα−Cβ−Sγ dihedral angle (χ-1)	1.40	3
Cα−Cβ−Sγ−Sγ dihedral angle (χ-2)	0.70	3
Cβ−Sγ−Sγ−Cβ dihedral angle (χ-3)	4.00	2

angle as used above, this distance can be calculated to be 3.03 Å. It is evaluated while rotating Sγ2 about the Cα2−Cβ2 bond. In the ideal case Sγ2 thus lies on a sphere with Cβ1 as the center and a radius of 3.03 Å. At the same time Sγ2 must lie on a circle formed when rotating Sγ2 about the Cα2−Cβ2 bond. This gives us 0, 1 or 2 possible solutions for the angle of rotation about the Cα2−Cβ2 bond, at which sphere and circle intersect. However, if the Cβ atom of the first residue is approximately in line with the Cα−Cβ bond of the second residue, this method will fail because all χ-1 angles in the second residue will be equally good. Therefore, if this is the case, we take the preferred values of −60°, +60° and 180° for this χ-1 angle. The procedure is repeated for the χ-1 angles of the first Cys residue. Combination of the results can give up to nine possible SS-bond conformations. Usually, in these calculated conformations the Sγ1−Sγ2 distance is far from ideal. Therefore, to minimize the deviations from perfect geometry, an energy minimization procedure is applied.

Energy refinement of calculated conformations

To get the best compromise between bond stretching energies, bond bending energies and torsion angle energies, we applied a simple energy minimization procedure to the atoms in a given disulfide conformation. Formulae and parameters to calculate energies were taken from the GROMOS molecular dynamics and energy refinement program package (Van Gunsteren and Berendsen, 1987) (see Table I for details). The actual energy refinement was performed with the algorithm of Dodson *et al.* (1976), which was modified to allow restraints on the torsion angles. Weights used in the minimization were adapted to the values used in the calculation of the energies (Table I).

Test data

To test the program a database consisting of 36 highly refined high resolution protein structures was used. These structures are listed in Table II together with some pertinent information on resolution and *R*-factor. Most of these structures were taken from the Brookhaven protein data bank (Bernstein *et al.*, 1977); a few were obtained separately from their authors (see Table II).

Results

Refined protein structures often differ considerably in the tightness of the restraints on bond lengths, bond angles and planarity of groups applied during the refinement. Therefore, we analyzed the agreement of the generated Cβ positions and the (observed) Cβ positions found in our database. The overall r.m.s. difference was 0.160 Å. This r.m.s. difference ranges from 0.101 to 0.321 Å (see Table II). We use this r.m.s. difference as a rough estimate of how well the model under investigation conforms to the ideal bond lengths and bond angles used in the program.

Selection of acceptable disulfide bonds

On the basis of the Cβ−Cβ distance. As mentioned under Materials and methods, the χ-3 angle in a disulfide bond has a strong preference for values of ±90°. Associated with this dihedral angle is a Cβ−Cβ distance of 3.83 Å. Indeed, in our database the average Cβ−Cβ distance is 3.83 ± 0.18 Å for Cys residues involved in a disulfide bond, with a range from 3.45 to 4.5 Å (corresponding to a χ-3 angle of −68° and +159°, respectively). The disulfide bond from residue 168 to 182 in rat mast cell protease has χ-3 = 159°. The χ-3 angle of the same SS-bond in a second molecule which is related to the first molecule by local symmetry, has a more normal value of 116°. If we leave out this one exceptional case, all other Cβ−Cβ distances fall within 3.83 ± 4 times the r.m.s. difference between observed and calculated Cβ positions. Therefore, this range for the Cβ−Cβ distance in a disulfide bridge is built into the program. Optionally, an angle increment can be specified, by which the χ-3 angle can deviate from the preferred angle of ±90°, in addition to what is allowed by four times the r.m.s. difference between observed and calculated Cβ positions.

On the basis of the energy of the predicted disulfide bond. As described under Materials and methods up to nine different conformations for a disulfide bond may be generated. After energy minimization these SS-bond conformations are ranked according to their energy. Input parameters of the program allow the user to select how many conformations are printed. By default, conformations with an energy < 10 kcal/mol are selected, unless a conformation has already been printed with an energy > 5.0 kcal/mol lower than that of the current conformation. This allows one to select only the energetically most favorable disulfide bonds. In our database the energies of the observed SS-bonds are mostly between 2 and 6 kcal/mol (average 4.1 ± 2.4 kcal/mol), with four SS-bonds having an energy > 10 kcal/mol. These are the 138−161 SS—bond in carboxypeptidase (12.8 kcal/mol; caused by a rather short Sγ1−Sγ2 bond of 1.89 Å), the 168−182 SS-bonds in both molecules of rat mast cell protease (16.5, respectively 10.6 kcal/mol; predominantly caused by high energy torsion and bond angles), and the 56−95 SS-bond in papain (10.35 kcal/mol).

Tests of the algorithm

To ascertain the correct working of the program we applied it to our test data to see if the program correctly predicts the conformations of disulfide bridges which are already present in the model. This is indeed the case: the r.m.s. difference between observed and predicted torsion angles in the disulfide bond is 6.6 ± 6.1°, 6.9 ± 7.2° and 8.0 ± 5.5° for χ-1, χ-2 and χ-3 respectively. The largest differences between observed and predicted SS-bonds were found in carboxypeptidase (the 138−161 SS-bond) and rat mast cell protease (the 168−182 SS-bonds). As mentioned above, these SS-bonds have unrealistically

Table II. Protein coordinate sets used for the database

Protein code name[a]	Description of protein	Resolution (Å)	*R*-factor	R.m.s. difference between observed and calculated Cβ positions
1BP2	Bovine phospholipase A_2	1.70	0.171	0.157
1CRN	Crambin	1.50	0.104	0.102
1CTF	L7/L12 ribosomal protein	1.70	0.174	0.137
1FB4	Immunoglobulin FAB-Kol	1.90	0.189	0.123
1GP1	Glutathione peroxidase	2.00	0.171	0.121
1HMQ	Hemerythrin	2.00	0.173	0.129
1ICB	Calcium-binding protein	2.00	?	0.215
1INS	Insulin	1.50	0.179	0.178
1LZ1	Human lysozyme	1.50	0.177	0.109
1MBD	Oxy-myoglobin	1.40	?	0.189
1PCY	Plastocyanin	1.60	0.17	0.120
1SN3	Scorpion neurotoxin	1.80	0.16	0.143
1TPP	β-Trypsin	1.40	0.191	0.135
2ALP	α-Lytic protease	1.70	0.131	0.125
2APP	Penicillopepsin	1.80	0.136	0.121
2AZA	Azurin	1.80	0.157	0.161
2CAB	Carbonic anhydrase	2.00	0.193	0.133
2CCY	Cytochrome c′	1.67	0.188	0.211
2CDV	Cytochrome c3	1.80	0.176	0.157
2LZM	T4 lysozyme	1.70	0.193	0.177
2OVO	Ovomucoid inhibitor	1.50	0.199	0.123
2RHE	Rhe Bence Jones protein	1.60	0.149	0.190
2SGA	*Streptococcus griseus* protease A	1.50	0.126	0.122
3C2C	Cytochrome c2	1.68	0.175	0.321
3CTS	Citrate synthase	1.70	0.192	0.141
3RP2	Rat mast cell protease	1.90	0.191	0.160
3WGA	Wheat germ agglutinin	1.80	0.179	0.158
4CYT	Cytochrome c	1.50	0.173	0.101
4DFR	Dihydrofolate reductase	1.70	0.155	0.262
4FXN	Flavodoxin	1.80	0.200	0.127
5CPA	Carboxypeptidase	1.54	0.190	0.153
5PTI	Bovine trypsin inhibitor	1.00	0.200	0.194
5RSA	RNase	2.00	0.159	0.189
5RXN	Rubredoxin	1.20	0.115	0.137
[b]	Papain	1.65	0.161	0.152
[c]	Subtilisin–eglin	1.20	0.18	0.148

[a]Protein code name as present in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).
[b]Coordinates as described by Kamphuis *et al.* (1984).
[c]Coordinates as described by Bode *et al.* (1987).

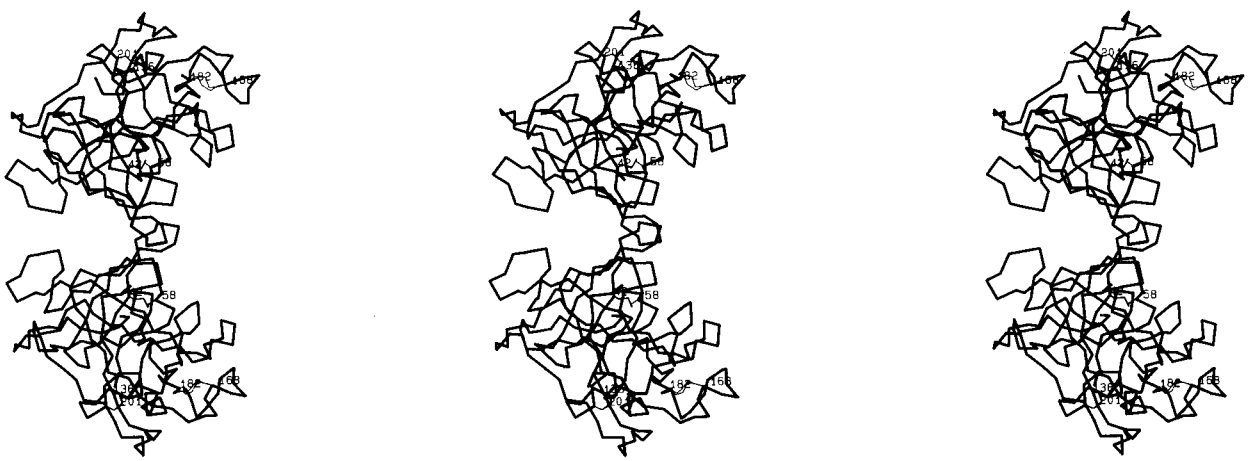

Fig. 5. Cα-Carbon tracing of the rat mast cell protease dimer. Indicated are the observed and predicted disulfide bond conformations. The program FRODO (Jones, 1985) has been used for this and the following figure. The three-picture stereo system used in this figure enables readers with both normal and cross-over stereo vision to view the images. For normal vision, select the left and centre images; for cross-over vision, use the centre and right images.

high energies for the observed conformations, and this suggests that the discrepancies between observed and predicted conformations are caused by the observed conformations rather than by the predicted ones. The other disulfide bonds are predicted correctly. To illustrate this, Figure 5 shows the rat mast cell protease dimer with the observed and predicted disulfide conformations, and Table III gives the numerical value of the torsion angles in this protein. It is clear from this example of one of the worst cases, that the predictions do agree with the observed conformations.

A more rigorous test is to compare predictions based on a wild-type structure with the conformations observed in a mutant protein into which SS-bonds have been engineered. Figure 6 shows the three-dimensional structure of subtilisin Carlsberg (Bode *et al.*, 1987), in which all predicted SS-bonds with an energy of <3 kcal/mol are indicated. Table IV lists all predicted SS-bond conformations with an energy <4 kcal/mol. Although in subtilisin BPN′ SS-bridges have been introduced and not in the subtilisin Carlsberg, a comparison can still be made because the folding of the two subtilisins is very similar. The structure of two SS-bonds introduced in subtilisin BPN′ has been determined (Katz and Kossiakoff, 1986). The χ-1-1, χ-2-1, χ-3, χ-2-2 and χ-1-2 angles were 42, 121, −98, 143 and −49° for the 22−87 SS-bond and −65, −50, 96, −171 and −157° for the 24−87 SS-bond. Comparison with the values given in Table IV shows a good agreement, although the observed 22−87 SS-bond is not the conformation for which we predict the lowest energy. Nevertheless, under the assumption of only small changes in the main chain conformation, we can reliably predict disulfide bond conformations.

A second protein into which a disulfide bond has been engineered is dihydrofolate reductase (Villafranca *et al.*, 1983, 1987). In contrast to our results with subtilisin, our predictions for the 39−85 SS-bond in the dimeric dihydrofolate reductase are quite different from the observed conformations. While the observed disulfide bonds have dihedral angles of ~ −150, −155, −80, −65 and −80, we predict two conformations of about equal energy: 60, −145, −90, 120, 180 and 60, 140, 105, −80, −105 for χ-1-1, χ-2-1, χ-3, χ-2-2 and χ-1-2 respectively. The reason for this discrepancy is that the main chain conformations of residues 39 and 85 have to change substantially in order to accommodate this SS-bond. Without any main chain readjustments the two Sγ atoms would have been 3.45 Å apart at the χ-1 torsion angles given by Villafranca *et al.* (1987). This directly shows the limitations of our approach, which assumes that main chain conformations stay more or less the same.

Computer time requirements

For a medium-sized protein like subtilisin (274 residues), the program takes ~5 min of CPU time on a microVAX-II computer. The CPU time needed is a function not only of the number of amino acid residues present in the model, but also of which values for the χ-3 angle are allowed. E.g. in the highly refined subtilisin of Bode *et al.* (1987), allowing a χ-3 range of $\pm(90 \pm 51)°$, 77 potential sites for an SS-bond are selected for which possible conformations will be analyzed. This results in 47 potential SS-bond sites with a conformational energy of ≤ 5 kcal/mol (CPU time 286 s). Extending the χ-3 range to $\pm(90.0 \pm 90.0)°$ leads to 182 sites to be analyzed. The full analysis in this case takes 495 s.

Table III. Comparison of observed and predicted dihedral angles in rat mast cell protease

Residues		χ1-1	χ2-1	χ3	χ2-2	χ1-2	Energy
42− 58	Observed	−81.2	−153.3	−88.2	−86.2	−70.9	3.34
	Predicted	−78.2	−141.0	−89.9	−87.6	−74.4	2.85
136−201	Observed	−57.9	−125.8	113.3	−79.6	−50.7	3.76
	Predicted	−56.2	−130.7	112.4	−79.1	−51.3	3.07
168−182	Observed	−51.6	82.4	158.9	−50.1	−92.0	16.50
	Predicted	−90.1	90.9	72.2	88.7	−165.2	5.22
42− 58	Observed	−84.4	−151.5	−80.5	−86.8	−66.8	5.01
	Predicted	−96.3	−144.4	−79.8	−83.1	−65.5	3.54
136−201	Observed	−49.9	−140.0	98.9	−68.8	−59.7	4.02
	Predicted	−53.1	−133.9	104.3	−79.1	−53.7	2.32
168−182	Observed	−61.5	76.7	116.0	48.1	−160.6	10.58
	Predicted	−84.3	91.0	82.5	80.9	−177.3	2.44

Discussion

As an aid in the selection of sites in a protein where a disulfide bridge might be introduced, we have developed a computer program to select those sites and to analyse the possible disulfide conformations. From the few examples available at this moment it is clear that under the assumption of no, or at the most, small conformational changes in the main chain conformation, the

Fig. 6. Cα-Carbon tracing of subtilisin Carlsberg with predicted SS-bond conformations, which have an energy of <3.0 kcal/mol. If more than one conformation with an energy <3.0 kcal/mol has been predicted, the conformation with the lowest energy is depicted. The three-picture stereo system used in this figure enables readers with both normal and cross-over stereo vision to view the images. For normal vision, select the left and centre images; for cross-over vision, use the centre and right images.

Table IV. Predicted SS-bonds in subtilisin Carlsberg with an energy of ≤4.0 kcal/mol

Residues	χ1-1	χ2-1	χ3	χ2-2	χ1-2	Energy
12–271	−160.8	−89.6	−94.0	134.2	−62.3	2.8
	80.6	161.3	−110.4	−62.7	32.1	3.6
15–271	70.7	−170.6	−106.4	−73.3	−28.4	2.8
22– 87	64.7	64.5	101.1	−137.0	62.1	1.6
	167.1	−136.7	101.0	64.2	−40.6	2.5
	64.0	127.9	−106.7	131.1	−43.6	4.0
23– 88	−167.9	−156.6	109.6	−56.4	−50.7	2.8
	72.1	155.6	−94.1	−23.4	−25.3	3.7
23–232	−65.9	−144.5	71.7	64.0	78.2	2.6
	166.7	65.9	72.3	−136.8	−161.5	3.9
24– 87	−64.2	−44.7	116.8	−155.4	171.9	2.7
	−40.4	−9.9	−105.7	−153.8	68.7	3.4
29–114	66.8	−142.2	−104.9	150.4	−167.2	3.9
29–119	31.4	−74.0	−114.5	−174.6	−156.8	3.9
32– 65	166.5	−72.1	−90.6	−34.4	−58.3	1.5
33– 65	62.8	29.0	−104.4	−61.3	27.0	3.0
	54.4	40.8	−109.8	−62.5	20.3	3.5
34– 60	148.2	−46.1	−90.5	−16.3	−44.6	3.6
35– 92	69.0	−141.2	90.2	−161.2	−44.8	2.0
	−31.0	60.1	106.4	83.1	−152.3	3.8
47– 57	179.8	−152.4	70.4	169.2	−177.0	2.3
50–106	74.2	66.7	114.6	−40.3	−156.4	3.8
57– 92	−155.8	−58.2	−116.5	−61.1	151.8	4.0
71–225	30.0	134.6	−97.6	162.4	45.0	3.6
73– 83	−82.0	43.8	−101.2	161.8	173.8	2.4
76– 86	26.2	−51.2	−91.3	−147.8	162.2	3.3
111–138	61.0	98.3	117.2	−154.8	−166.2	3.6
114–119	158.7	2.9	107.5	169.6	46.0	3.7
123–228	−138.5	62.0	91.6	157.5	−56.9	3.0
127–166	−164.4	−67.3	−99.3	−69.1	144.3	3.1
151–169	174.0	−79.8	−86.6	12.3	−67.1	3.0
154–166	−38.1	−75.1	−109.2	−43.7	42.4	2.9
154–191	17.5	48.3	75.3	179.7	62.9	3.2
163–193	172.6	85.4	−81.9	138.5	177.1	2.0
	−100.0	−93.2	76.8	70.3	175.8	3.5
179–187	164.7	−61.1	−91.7	168.6	−27.9	2.3
181–203	−60.8	−27.5	−96.7	163.1	−7.1	3.9
205–222	56.6	158.5	102.4	−173.3	−136.5	3.7
209–215	−44.6	−158.3	−81.2	−61.3	70.4	1.4
	−30.5	−144.5	−99.4	−56.9	58.4	2.6
	37.5	157.0	102.8	56.6	−34.4	2.9
269–272	48.0	65.0	92.8	−149.3	−68.7	2.1
	161.0	−151.2	84.4	71.1	−42.1	2.8

program is capable of correctly predicting disulfide conformations.

The program described in this paper appears to be faster than the PROTEUS program written by Pabo and Suchanek (1986). This latter program was reported to take 15 min of CPU time on a VAX 11/750 computer for the N-terminal domain of λ repressor (2 × 92 residues).

Our program calculates the conformations with minimal energies for potential disulfide bonds at given sites. It does so, however, without regard for neighboring atoms, which might cause Van der Waals collisions. Neither are packing defects taken into account, which might arise from the substitution of amino acid sidechains by cysteine sidechains. Therefore, after selecting the most promising conformations, one should examine these conformations for too close contacts on a computer graphics device, and perform an energy minimization of the whole protein molecule to remove unfavorable interactions and packing defects.

Acknowledgements

We thank Dr W.Bode for making available the nagarse–eglin coordinates prior to publication. We thank Profs Jan Drenth and Wim Hol for stimulating discussions.

References

Altman,S.L. (1986) *Rotations, Quaternions and Double Groups.* Clarendon Press, Oxford, pp. 73–75.
Anfinsen,C.G. and Scheraga,H.A. (1975) *Adv. Protein Chem.*, **29**, 205–299.
Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **141**, 441–484.
Bode,W., Papamokos,E. and Musil,D. (1987) *Eur. J. Biochem.*, **166**, 673–692.
Dodson,E.J., Isaacs,N.W. and Rollett,J.S. (1976) *Acta Crystallogr.*, **A32**, 311–315.
Jones,T.A. (1985) *Methods Enzymol.*, **115**, 157–171.
Kamphuis,I.G, Kalk,K.H., Swarte,M.B.A. and Drenth,J. (1984) *J. Mol. Biol.*, **179**, 233–256.
Katz,B.A. and Kossiakoff,A. (1986) *J. Biol. Chem.*, **261**, 15480–15485.
Pabo,C.O. and Suchanek,E.G. (1986) *Biochemistry*, **25**, 5987–5991.
Pantoliano,M.W., Ladner,R.C., Bryan,P.N., Rollence,M.L., Wood,J.F. and Poulos,T.L. (1987) *Biochemistry*, **26**, 2077–2082.
Perry,L.J. and Wetzel,R. (1984) *Science*, **226**, 555–557.
Richardson,J. (1981) *Adv. Protein Chem.*, **34**, 167–339.
van Gunsteren,W.F. and Berendsen,H.J.C. (1987) BIOMOS. Biomolecular Software, Laboratory of Physical Chemistry, University of Groningen, Groningen, The Netherlands.
Villafranca,J.E., Howell,E.E., Voet,D.H., Strobel,M.S., Ogden,R.C., Abelson,J.N. and Kraut,J. (1983) *Science*, **222**, 782–788.
Villafranca,J.E., Howell,E.E., Oatley,S.J., Xuong,N. and Kraut,J. (1987) *Biochemistry*, **26**, 2182–2189.
Wells,J.A. and Powers,D.B. (1986) *J. Biol. Chem.*, **261**, 6564–6570.
Wetzel,R. (1987) *Trends Biochem. Sci.*, **12**, 478–482.

Received on March 23, 1988; revised on May 17, 1988